Title: Tracking type specific prevalence of human papillomavirus in cervical pre-cancer: a novel sampling strategy

Authors:

Edward K Waters (ewaters@kirby.unsw.edu.au)
John Kaldor (jkaldor@kirby.unsw.edu.au)
Andrew J Hamilton (andrewih@unimelb.edu.au)
Anthony MA Smith (anthony.smith@latrobe.edu.au)
David J Philp (dphilp@kirby.unsw.edu.au)
Basil Donovan (bdonovan@kirby.unsw.edu.au)
David G Regan (dregan@kirby.unsw.edu.au)

Version: 2 Date: 31 January 2012

Author's response to reviews: see over
Concerning the manuscript

**Tracking type specific prevalence of human papillomavirus in cervical pre-cancer: a novel sampling strategy**

Edward K Waters, John Kaldor, Andrew J Hamilton, Anthony MA Smith, David J Philp, Basil Donovan, David G Regan

We thank the editors for the recent reception of the reviews of this paper. We are gratified that the second reviewer, Dr Antonella Agodi, viewed the manuscript as “an article of importance in its field”. We note however that Dr Agodi suggested that the manuscript be reviewed by a statistician regarding potential methodological issues. In his report, the second other reviewer Dr Mark Myatt notes he has provided a statistical review. In his careful review Dr Myatt suggested a number of changes that would improve the manuscript. We have responded to each of methodological and stylistic points he raised in detail below.

Abstract : Paragraph 2 : Method : The use of the term "precision" is wrong since the reported method classifies rather than estimates prevalence. I think that the authors mean "accuracy" or, perhaps, "performance" or "utility".
This point is correct and we have changed instances of “precision” to “accuracy”.

Abstract : Paragraph 3 : Results : The use of “false negative” and “false positive” is confusing since a three-class method is being tested. This reflects a confusion throughout the report. It is not clear whether a two or three class method is required and which is being tested.

Although one can classify the incorrect acceptance or rejection of any hypothesis using the terms “false negative” and “false positive”, we accept that when multiple hypotheses are being tested as in this case (a three-class method) the use of these terms could lead to confusion unless it is clearly specified which of the three classifications has been incorrectly accepted or rejected in every instance. We therefore have refined the manuscript using either more general or more appropriate terminology, as required for greater clarity.

Page 5 : "... in public health" : Some references to public health applications of sequential sampling methods is required. These methods (commonly called “LQAS”) are very common in service epidemiology.
Page 5 : "... method proposed ..." : The truncated sequential sampling method has been adopted by the WHO and CDC and is in current use in many settings.

We have altered the manuscript accordingly.

Page 5 : "... prevalence estimate ..." : Do the authors mean “classification”?

Although a classification of low, medium or high prevalence also implies that the prevalence is estimated to be within a particular range of values, we have altered the manuscript to use the term classification to reduce any ambiguity.

Page 5 : "... false positive or a false negative" : See above.
Page 5 and elsewhere : The text seems to switch between reporting a two class method and a three class method.
page 5 : "... estimate of prevalence" : See above.
Page 5 : "... high precision" : See above.

As noted above we have incorporated suggestions regarding the terms precision, false positive and false negative. Changing the terminology assists in clarifying that a three class method is being tested.

Page 6: Formula 1 : This formula seems to be incorrectly specified. The numerator should be:
-ln((1 - alpha) / beta))

The formulae that we have used have been in use for decades used by authorities in the field, particularly in the field of applied entomology. We can provide many citations, but we suggest that it is sufficient that the editors check the first and second authors’ own paper (reference 12 in the manuscript) and a book which is cited by Dr Myatt and in his own work, being:
Binns, Nyrop, and van der Werf: Sampling and monitoring for crop protection: the theoretical basis for developing practical decision guides. Oxford, CABI, 2000 (We have supplied the editors with a scan of relevant pages of this book for ease of reference).

Both suggested references give the formula in the same way we have specified it.

Nevertheless, Furthermore, we have examined the value of the substitution suggested by Myatt. The substitution suggested by Myatt is similar in form to the numerator for the formula giving the lower intercept, but formula 1 refers to the upper intercept; we therefore wonder if Myatt intends that the substitution should be made in the equation for the lower intercept (formula 2) instead of in Formula 1. We respond to both possibilities here.

If the substitution is intended to be made in Formula 1 as written, it is incorrect because it gives an upper intercept which is identical in value to the lower intercept (upper intercept = -1.56, lower intercept = -1.56), which is not acceptable for these purposes.

If the substitution is actually intended for the lower intercept (Formula 2 in our paper), the lower intercept calculated as given in our paper is actually equal in value to the lower intercept calculated by Myatt’s method (both equal to -1.56), because the equations are mathematically equivalent. Therefore, we have chosen not to implement this substitution since it will either not make a difference if intended for the lower intercept or give an inappropriate result if intended for the upper. We attach a spreadsheet showing our calculations.

Page 6 : Formula 3 : The numerator might be simplified to:
\[ \ln(q1 / q2) \]

We have altered formula 3 as suggested.

Page 6 : Paragraph 1 : The text is referring to a two class approach.

Upon examination of the manuscript by Dr Myatt in which the truncated sampling method (reference 17 in the manuscript) we cannot see how the formulae for the upper and lower intercepts and slope differ from Wald’s formulae for a two class method, save for the use of the negative log of the inverse numerator, which is mathematically the same as the log of the numerator. We therefore believe it is accurate to attribute the first three formulae to Wald.

However we found that the formula for the upper intercept (formula 3 in Myatt and Bennett’s paper) is mis-specified, as it gives an upper intercept which is lower than the lower intercept (upper = -2.77, lower = -1.56 in the spreadsheet that we have attached). We presume this reflects a typographical error.
The existence of a third classification has nothing to do with the solution of these formula, but is rather given by the implementation of truncation based on the solution of formula 4 in both ref. 17 and in the manuscript; therefore up until the truncation point the three-class truncated sampling method provides results that are identical to a two class method with a minimum sample size imposed.

We also found that the formula for the upper intercept (formula 3 in Myatt and Bennett’s paper) is mis-specified, as it gives an upper intercept which is lower than the lower intercept (upper = 2.77, lower = 1.56 in the spreadsheet that we have attached).

Page 6 : Paragraph 1 : It is unclear what is intended by the "... plus tolerance boundaries".
In entomology, it is common to impose tolerance bounds of say 0.05 around thresholds to express uncertainty, so for example an upper threshold of 0.15 would become 0.2 (see the authors’ paper – ref. 12 in the manuscript). We note however that Myatt did not use this approach in his paper, so on further reflection we have chosen to remove these tolerance bounds around our thresholds as they appear to be a source of confusion.

Page 6 : Paragraph 1 : q2 should be defined as 1 - p2 (not just "p2").
We thought that this was implied by the context, but have changed the text to remove any ambiguity.

Page 6 : Formula 4 : This is the formula for a two-class binomial sequential sampling classifier. It seems that the authors are confused between the two-class binary sequential sampling method and the three class TSS method.

The confusion arises because in Myatt’s original manuscript in Figure 3D he truncates sampling at Nmax, and only describes later in his manuscript that this is a different Nmax to that given by Formula 4, and is in fact selected arbitrarily.

Page 6 : Paragraph 3 : It is unclear what is intended by the "... plus tolerance boundaries".
See above.

Page 7 : Paragraph 3 : Substituting:
p1 = 0.05 <- from Page 7 : Paragraph 2
p2 = 0.15 <- from Page 7 : Paragraph 2
alpha = 0.01 <- from Page 7 : Paragraph 2
beta = 0.025 <- from Page 7 : Paragraph 2
yields N_max = 138 not the 48 quoted in the text. This (large) discrepancy should be accounted for.

Performing the same calculations as Myatt, we get 130 rather than 138; however, the discrepancy in the original manuscript existed because of a typographical error; in fact Nmax was chosen arbitrarily, as done in Myatt’s original paper where values of 25-60 were chosen. We have therefore decided to adopt a similar approach and not use this formula.
Page 8 : Paragraph 1 : "... uniform distribution ..." : Myatt and Bennett did not use a uniform distribution of prevalences but a distribution reflecting reported prevalences of HIVDR (see Table 3 in Myatt and Bennett). Uniform distribution within band was used but the frequency of each band was non-uniform. The presentation of the use of a triangular distribution as "... a methodological enhancement intended to simulate the most frequent sampled value for the prevalence ..." is not, therefore, correct.
We apologise for misinterpreting Myatt’s paper, but because we are not aware of other authors who have used triangular distributions or compared the performance of a sampling plan under different distributions of positivity this may still represent an interesting methodological enhancement.

Page 8 : Paragraph 2 : The fitting of separate probability of classification curves is a methodologically flawed approach. The three curves are not independent of each other. For example, if \( P(\text{Low}) = 0.2 \) and \( P(\text{High}) = 0.1 \) the \( P(\text{Moderate}) \) is constrained to be 0.7. A better approach would to have been to use a larger number of simulation replicates in at each of the tested prevalences and plot the calculate probability of classification.
We agree with this and think that the suggested approach is a significant improvement – we have implemented it in the text. We have greatly increased the number of simulations (10,000 resampling bouts on 1000 prevalences) to facilitate the production of the suggested figure.

Page 8 : Paragraph 4 / Figure 2 : It is unclear whether prevalences in excess of 30% are likely to be found in real-world applications of this method. If such prevalences are unlikely then testing at prevalences above 30% makes little sense.
We have retained the current scheme, since in CIN 3 types such as HPV 16 can be responsible for large numbers of lesions (up to at least 95% in some small samples). An additional sentence has been added explaining this point.

Page 8 : Paragraph 4 : The \( R^2 \) values are without clear use or meaning here. Because we have discarded the fitting of curves as noted above, this point is now not relevant.

Page 9 : Paragraph 2 / Figure 2 : Figure 2 looks wrong:
(1) In the sense that it represents poor performance of the classifier at returning a moderate prevalence classification. At \( N_{\text{max}} = 47 \) the HIVDR classifier works better than this \( N_{\text{max}} = 48 \) classifier.
(2) It seems that the probability of classification varies considerably between prevalences that are just 1% apart. This plot looks more like 100 prevalences sampled 50 times each (i.e. 5000 replicates) was used rather than 100 prevalences sampled 5000 times each (i.e. 500 000 replicates was used). The authors should check their computer code.
Table 1 : The average sample numbers appear to be incorrect. The average sample number should be much higher for moderate prevalence classifications. In adopting Myatt’s suggested approach to generating Figure 2 (see above), we have now increased the number of simulations (10,000 replicates over 1000 prevalences) to produce a better graph, so these comments are now inapplicable, as
both Figure 2 and Table 1 have been updated on the basis of new simulations, and the text updated where appropriate. We have attached the computer code used to assist further review.

**Concluding remarks.**

We would especially like to thank Dr Myatt for his careful review, which we feel has improved the paper immeasurably in many respects. We have therefore included Dr Myatt in our acknowledgements section for his helpful feedback, and thank BMC Medical Research Methodology for the opportunity to respond to these comments and to revise the paper.

Yours sincerely

Edward K Waters